

(12)特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2003 年 12 月 24 日 (24.12.2003)

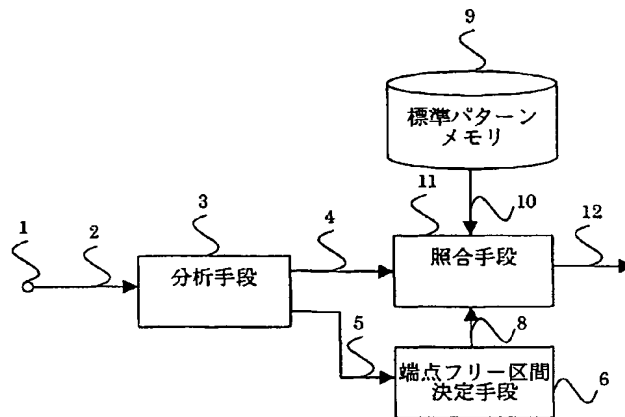
PCT

(10) 国際公開番号  
WO 03/107326 A1

- (51) 国際特許分類: G10L 11/02 千代田区 丸の内二丁目 2 番 3 号 三菱電機株式会社  
Tokyo (JP).
- (21) 国際出願番号: PCT/JP02/05847
- (22) 国際出願日: 2002 年 6 月 12 日 (12.06.2002)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人 (米国を除く全ての指定国について): 三菱電機株式会社 (MITSUBISHI DENKI KABUSHIKI KAISHA) [JP/JP]; 〒100-8310 東京都千代田区丸の内二丁目 2 番 3 号 Tokyo (JP).
- (74) 代理人: 宮田 金雄, 外 (MIYATA, Kaneo et al.); 〒100-8310 東京都千代田区丸の内二丁目 2 番 3 号 三菱電機株式会社内 Tokyo (JP).
- (81) 指定国 (国内): CN, JP, US.
- (84) 指定国 (広域): ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- 添付公開書類:  
— 国際調査報告書
- (72) 発明者: および
- (75) 発明者/出願人 (米国についてのみ): 花沢 利行 (HANAZAWA, Toshiyuki) [JP/JP]; 〒100-8310 東京都
- 2 文字コード及び他の略語については、定期発行される各 PCT ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

(54) Title: SPEECH RECOGNIZING METHOD AND DEVICE THEREOF

(54) 発明の名称: 音声認識方法及びその装置



3...ANALYZING MEANS

9...STANDARD PATTERN MEMORY

11...COLLATING MEANS

6...ENDPOINT-NODE FREE SECTION DETERMINING MEANS

(57) Abstract: A speech recognizing method and a device thereof capable of recognizing speech with endpoint-node free, and improving the recognition performance under noise. This speech recognizing method comprises an analysis step of performing the sound analysis of the input speech, an endpoint-node free section determining step of detecting a section in which the power of the input speech continuously exceeds a predetermined threshold as a peak section, and outputting combination of the start-point free section with the endpoint free section by assuming that the peak section with the maximum power as the maximum

(続葉有)

WO 03/107326 A1

BEST AVAILABLE COPY



---

peak section, and a start-point free section is present before the point at which the maximum peak section is below the threshold, and the endpoint free section is present after the point at which the maximum peak section exceeds the threshold, and a collating step of performing the pattern matching of the pattern with the standard pattern in each of the combinations.

(57) 要約:

端点フリーによる音声認識を行いかつ騒音下の認識性能を改善できる音声認識方法とその装置を提供する。

入力音声を音響分析する分析ステップと、入力音声のパワーが所定の閾値を連続して超える区間を山区間として検出し、パワーが最大となる山区間を最大山区間として、その最大山区間が前記閾値を下回る点以前に始端フリー区間が存在すると仮定し、前記最大山区間が前記閾値を上回る点以降に終端フリー区間が存在すると仮定して始端フリー区間と終端フリー区間の組み合わせを出力する端点フリー区間決定ステップと、前記組み合わせのそれぞれのパターンと標準パターンとのパターンマッチングを行う照合ステップとを備えた。

## 明 細 書

## 音声認識方法及びその装置

## 技術分野

- 5 本発明は、騒音下の認識性能を改善する音声認識方法及びその装置に関するものである。

## 背景技術

- 10 音声認識は、入力信号のパワーを予め設定した閾値と比較して閾値以上の区間を音声区間として検出し、予め用意してある標準パターンとパターンマッチングを行うことによってなされる。

- しかし騒音下では、騒音のパワーレベルが前記閾値を越えることがあるため、音声区間検出を誤ることとなり、誤認識の原因となるという問題点があった。そのための対策として、音声区間の始端と終端を一意に定めず、始端と終端の取り得る組み合わせの全ての区間に対してパターンマッチングを行う端点フリーと呼ばれる認識方法がある。しかし全ての区間に対してパターンマッチングを行うと「東大阪(ひがしおおさか)」という発声が「大阪(おおさか)」と
- 15 認識される等の部分マッチングによる誤認識が問題となる。

- これを低減するため、従来から始端と終端の取り得る組み合わせの範囲を制限する、すなわち端点フリーの範囲を制限する技術がある。この技術の一例として特開昭63-300295がある。
- 25

図5は、特開昭63-300295記載の音声認識装置

の一構成例を示すブロック図である。本例では、特定話者の単語認識を行う場合を説明する。信号の入力端 1 から入力信号 2 を入力すると、分析手段 3 は入力信号 2 を A/D 変換し、フレームと呼ぶ短い時間区間ごとに分割して、各フレームごとに音響分析を行い、各フレームごとに特徴ベクトル  $X(t)$  ( $t=1,2,3,\dots,T$ ) の時系列 4、およびパワー  $P(t)$  と零交差回数  $Z(t)$  ( $t=1,2,3,\dots,T$ ) の時系列 5 を計算して出力する。ここで  $T$  は、入力信号 2 の全フレーム数である。特徴ベクトル  $X(t)$  は、例えば LPC 分析によって得られる LPC ケプストラムとする。零交差回数  $Z(t)$  は、有声音区間の検出に用いる。すなわち有声音は低周波数成分にパワーが集中するため、零交差回数  $Z(t)$  は少ないので、本例では後述するように、音声パワーが所定値以上でかつ零交差回数  $Z(t)$  が少ない区間を有声音であるとみなしている。

15 有声音決定手段 7 は、入力信号のパワー  $P(t)$  と零交差回数  $Z(t)$  ( $t=1,2,3,\dots,T$ ) の時系列 5 とを入力として、図 6 に示すように、音声パワーが予め定めた閾値を上回るフレームを音声区間の始端 PB、閾値を下回るフレームを音声区間の終端端 PE として検出する。また検出した音声区間で、零交差回数  $Z(t)$  が予め定めた閾値未満となるフレームを有声音区間の始端 ZB、閾値以上となるフレームを有声音区間の終端 ZE として検出する。

そして、前記音声区間の始端 PB から有声音区間の始端 ZB までの区間を、始端フリー区間  $B_{free}=\{PB, ZB\}$  として決定する。ここで PB は、始端フリー区間の先頭フレーム、ZB は始端フリー区間の最終フレームである。また有声音

区間の終端 ZE から音声区間の終端 PE までの区間を、終端フリー区間  $E_{free}=\{ZE, PE\}$  として決定する。ここで ZE は、終端フリー区間の先頭フレーム、PE は終端フリー区間の最終フレームである。

- 5      有声音決定手段 7 は、前記始端フリー区間  $B_{free}=\{PB, ZB\}$  と終端フリー区間  $E_{free}=\{ZE, PE\}$  を端点フリー区間情報 8 として出力する。

- 照合手段 6 は、特徴ベクトルの時系列 4 と、端点フリー区間情報 8 である始端フリー区間  $B_{free}=\{PB, ZB\}$  と、終端フリー区間  $E_{free}=\{ZE, PE\}$  とを入力として、始端フリー区間と終端フリー区間内の全ての始終端の組み合わせで、標準パターンメモリ 9 に格納されている標準パターン 10 と、例えば DP マッチングによるパターンマッチングを行い、距離値が最小のものを認識結果とする。なお、標準パターンメモリ 9 には、予め認識対象とする各単語の標準パターン  $REF(i), (i=1, 2, 3, \dots, K)$  として、各単語の特徴ベクトルの時系列が格納されているものとする。ここで K は、標準パターンの数である。特徴ベクトル  $X(t)$  は、例えば LPC (Linear Predictive Coding) 分析によって得られる
- 15      LPC ケプストラムとする。

- 以上述べたように、従来技術では検出された音声区間と有声音区間の間に正しい音声区間の始終端が存在すると仮定して、端点フリー区間を制限している。しかし非定常騒音の種類は様々で、自動車内騒音のように、スペクトルの
- 25      低域にパワーが集中している騒音では零交差回数が少ないため、有声音区間と判定される危険性が大きい。騒音が有

声音と判定された場合には、騒音区間を含めてパターンマッチングを行うため、誤認識の原因となる。

- この発明は上記課題を解決をするためになされたもので、正確な判定が困難な有声音区間の判定を必要とせず、効率的に端点フリー区間を制限して、音声認識精度を改善する音声認識装置を提供することを目的とする。

#### 発明の開示

- 第 1 の発明に係る音声認識方法は、入力音声を音響分析しこの入力信号についてパワーを出力する分析ステップと、前記パワーが所定の閾値を連続して超える区間を山区間として検出し、パワーが最大となる山区間を最大山区間として、その最大山区間が前記閾値を下回る点以前に始端フリー区間が存在すると仮定し、前記最大山区間が前記閾値を上回る点以降に終端フリー区間が存在すると仮定して始端フリー区間と終端フリー区間の組み合わせを出力する端点フリー区間決定ステップと、前記組み合わせの始端フリー区間と終端フリー区間により特定されるパターンと標準パターンとのパターンマッチングを行う照合ステップとを備えたものである。

- 第 2 の発明に係る音声認識方法は、前記照合ステップが、前記端点フリー区間が出力するすべての始端フリー区間と終端フリー区間の組み合わせにより特定される各パターンと標準パターンとのパターンマッチングを行う構成とされたものである。

第 3 の発明に係る音声認識方法は、前記端点フリー区間

決定ステップが、検出された前記山区間のうち、パワーの累積が最大となる山区間を最大山区間とする構成とされたものである。

- 第4の発明に係る音声認識方法は、前記分析ステップが、
- 5 検出点毎にパワーを出力し、前記端点フリー区間決定手段が、検出された前記山区間のうち、前記検出点毎のパワーのうち上位の所定の個数の検出点のパワーの和が最大となる山区間を最大山区間とする構成とされたものである。

- 第5の発明に係る音声認識方法は、前記端点フリー区間
- 10 決定ステップが、前記最大山区間以前の山区間が前記閾値を上回る点の近傍に前記始端フリー区間が存在すると仮定し、前記最大山区間以後の山区間が前記閾値を下回る点の近傍に前記終端フリー区間が存在すると仮定する構成とされたものである。

- 15 第6の発明に係る音声認識方法は、前記端点フリー区間決定ステップが、前記始端フリー区間が存在すると仮定する範囲の幅と前記終端フリー区間が存在すると仮定する範囲の幅とを、各山区間毎に異なる構成とされたものである。

- 第7の発明に係る音声認識方法は、前記分析ステップが、
- 20 フレーム毎にパワーを出力し、前記端点フリー区間決定手段が、パワーが前記閾値を下回るフレームを前記閾値を下回る点とし、パワーが前記閾値を上回るフレームを前記閾値を上回る点とする構成とされたものである。

- 第8の発明に係る音声認識装置は、入力音声を音響分析
- 25 しこの入力信号についてパワーを出力する分析手段と、前記パワーが所定の閾値を連続して超える区間を山区間とし

- て検出し、パワーが最大となる山区間を最大山区間として、その最大山区間が前記閾値を下回る点以前に始端フリー区間が存在すると仮定し、前記最大山区間が前記閾値を上回る点以降に終端フリー区間が存在すると仮定して始端フリー区間と終端フリー区間の組み合わせを出力する端点フリー区間決定手段と、前記組み合わせの始端フリー区間と終端フリー区間により特定されるパターンと標準パターンとのパターンマッチングを行う照合手段とを備えたものである。
- 5 第 9 の発明に係る音声認識装置は、前記照合手段が、前記端点フリー区間が出力するすべての始端フリー区間と終端フリー区間の組み合わせにより特定される各パターンと標準パターンとのパターンマッチングを行う構成とされたものである。
- 10 第 10 の発明に係る音声認識装置は、前記端点フリー区間決定手段が、検出された前記山区間のうち、パワーの累積が最大となる山区間を最大山区間とする構成とされたものである。
- 15 第 11 の発明に係る音声認識装置は、前記分析手段が、検出点毎にパワーを出力し、前記端点フリー区間決定手段が、検出された前記山区間のうち、前記検出点毎のパワーのうち上位の所定の個数の検出点のパワーの和が最大となる山区間を最大山区間とする構成とされたものである。
- 20 第 12 の発明に係る音声認識装置は、前記端点フリー区間決定手段が、前記最大山区間以前の山区間が前記閾値を上回る点の近傍に前記始端フリー区間が存在すると仮定し、
- 25



前記最大山区間以後の山区間が前記閾値を下回る点の近傍に前記終端フリー区間が存在すると仮定する構成とされたものである。

第13の発明に係る音声認識装置は、前記端点フリー区間  
5 決定手段が、前記始端フリー区間が存在すると仮定する範囲の幅と前記終端フリー区間が存在すると仮定する範囲の幅とを、各山区間毎に異なる構成とされたものである。

第14の発明に係る音声認識装置は、前記分析手段が、フレーム毎にパワーを出力し、前記端点フリー区間決定手段  
10 が、パワーが前記閾値を下回るフレームを前記閾値を下回る点とし、パワーが前記閾値を上回るフレームを前記閾値を上回る点とする構成とされたものである。

#### 図面の簡単な説明

15 図1は本発明の実施例1と実施例2による音声認識装置の構成図、

図2は本発明の実施例1の始端フリー区間と終端フリー区間の決定方法の説明図、

20 図3は本発明の実施例2の始端フリー区間と終端フリー区間の決定方法の説明図、

図4は本発明の実施例2の始端フリー区間と終端フリー区間の決定方法の説明図、

図5は従来技術による音声認識装置の構成図、

25 図6は従来技術による始端フリー区間と終端フリー区間の決定方法の説明図である。

発明を実施するための最良の形態

実施例 1.

図 1 は、本発明に係る実施例 1 の音声認識装置の構成を示すブロック図である。同図において、1 は信号の入力を行う信号入力端、2 は信号入力端 1 によって入力された入力信号、3 は入力信号 2 について音響分析を行う分析手段、4 は分析手段 3 によって算出される入力信号の特徴ベクトルの時系列、5 は分析手段 3 によって算出される入力信号のパワーの時系列、6 は入力信号のパワーの時系列 5 に基づいて端点フリー区間を決定する端点フリー区間決定手段、8 は端点フリー区間決定手段 6 によって出力される端点フリー区間情報、9 は音声認識のための照合処理において用いられる標準パターンを記憶する標準パターンメモリ、10 は音声認識のための照合処理において用いられる標準パターン、11 は各単語の標準パターンとのパターンマッチングを行う照合手段である。

本実施例では、特定話者の単語認識を行う場合を説明する。標準パターンメモリ 9 には、予め認識対象とする各単語の標準パターン  $REF(i), (i=1, 2, 3, \dots, K)$  として、各単語の特徴ベクトルの時系列が格納されているものとする。ここで  $K$  は、標準パターンの数である。特徴ベクトル  $X(t)$  は、例えば LPC 分析によって得られる LPC ケプストラムとする。

次に動作について説明する。信号入力端 1 から入力信号 2 を入力すると、分析手段 3 は入力信号 2 を A/D 変換した後、短い時間区間ごとにフレームと呼ばれる単位に分割し

て、各フレームごとに音響分析を行ってそのフレームの特徴ベクトル  $X(t)$  ( $t=1,2,3,\dots,T$ )の時系列 4、およびパワー  $P(t)$  ( $t=1,2,3,\dots,T$ )の時系列 5 を計算して出力する。ここで、 $T$  は入力信号 2 の全フレーム数である。特徴ベクトル  $X(t)$  は、例えば LPC(線形予測)分析によって得られる LPC ケプストラムとする。パワー  $P(t)$  は、例えばフレーム内の入力信号のディジタル値を 2 乗和して対数をとることによって得られる。

端点フリー区間決定手段 6 は、入力信号のパワー  $P(t)$  ( $t=1,2,3,\dots,T$ )を入力として、図 2 に示すように、パワーが予め定めた閾値を越えたフレームが所定数以上連続する区間を、山区間  $H(i)=\{B(i), E(i)\}$  ( $i=1,2,3,\dots,N$ )として検出する。ここで、 $B(i)$  は山区間の始端、すなわちパワーが前記閾値を越えたフレームである。また、 $E(i)$  は山区間の終端、すなわちパワーが前記閾値未満となったフレームである。 $N$  は検出された山区間の個数である。図 2 の例では、 $N=3$  である。

次に、端点フリー区間決定手段 6 は、(1)式あるいは(2)式によって、各山区間  $PH(i)=\{B(i), E(i)\}$  ( $i=1,2,3,\dots,N$ )のパワー強度  $PK(i)$  ( $i=1,2,3,\dots,N$ )を計算し、(3)式にしたがってパワー強度が最大の山をパワー強度最大山区間  $PH(I)$ として検出する。(2)式において、 $t(i,j)$  は各山区間  $PH(i)=\{B(i), E(i)\}$  の区間内のフレームで、パワーが  $j$  番目に大きいフレームとする。したがって、 $t(i,1)$  は、山区間  $PH(i)$  内でパワーが最大のフレームとなる。このことより、(2)式は、各山区間について、当該区間内でパワーが大きい上位  $M$  個の

パワー値を用いてパワー強度を計算していることになる。  
 一方(1)式は、各山区間について、当該区間内の全パワー  
 の和としてパワー強度を計算している。(1)、(2)式のいず  
 れを用いるかについては、音声認識装置の使用時に想定さ  
 5 れる環境騒音の種類や認識対象とする音声によって、選択  
 すればよい。

$$PK(i) = \sum_{t=B(i)}^{t=E(i)} P(t), \quad (i=1, 2, 3, \dots, N) \quad (1)$$

$$PK(i) = \sum_{j=1}^M P(t(i, j)), \quad (i=1, 2, 3, \dots, N) \quad (2)$$

$$10 \quad I = \operatorname{argmax}(i) PK(i) \quad (3)$$

本実施例では、パワー強度最大山区間は音声であると仮  
 定するので、例えばパワーの山の継続時間は短いが、瞬間  
 的な信号パワーが大きくなるスパイク状の騒音では(1)式  
 15 を用い、逆にパワーの山の継続時間は長いが、フレーム毎  
 のパワーは大きくない騒音では(2)式を用いることにより、  
 騒音区間がパワー強度最大山区間とならないようにするこ  
 とができる。

次に端点フリー区間決定手段6は、(4)、(5)式にしたが  
 20 って最初の山区間の始端 B1 から、パワー強度最大山区間  
 の始端 B(I)までの区間に、始端マージン bm1, bm2を加え  
 た区間を、始端フリー区間 BF={bfL, bfR}として決定する。  
 ここで、bfL は始端フリー区間の先頭フレーム、bfR は始  
 端フリー区間の最終フレームであり、また始端マージン

bm1, bm2 は、予め定めた 0 以上の定数である。

$$bfL = B(I) - bm1 \quad (4)$$

$$bfR = B(I) + bm2 \quad (5)$$

5

次に(6)、(7)式にしたがって、端点フリー区間決定手段 6 は最後の山区間の終端 E(N)から、パワー強度最大山区間の始端 E(I)までの区間に、終端マージン em1, em2 を加えた区間を、終端フリー区間 EF={efL, efR}として決定する。

10 ここで、efL は終端フリー区間の先頭フレーム、efR は終端フリー区間の最終フレームである。終端マージン em1, em2 は、予め定めた 0 以上の定数である。以上の処理により決定された始端フリー区間 BF と終端フリー区間を図 2 に示す。

15

$$efL = E(I) - em1 \quad (6)$$

$$efR = E(N) + bm2 \quad (7)$$

端点フリー区間決定手段 6 は、前記始端フリー区間  
20 BF={bfL, bfR}と終端フリー区間 EF={efL, efR}を、端点フリー区間情報 8 として出力する。

照合手段 11 は、特徴ベクトルの時系列 4 と、端点フリー区間情報 8 による始端フリー区間 BF={bfL, bfR}と終端フリー区間 EF={efL, efR}とを入力として、始端フリー区  
25 間と終端フリー区間内の全ての始終端の組み合わせで標準パターンメモリ 9 に格納されている各単語の標準パターン

1 0 である REF(i)とパターンマッチングを行う。この処理  
を標準パターン REF(i) (i=1,2,3,...,K)の全てに対して順  
次行い、これら全ての始終端の組み合わせと標準パターン  
との照合処理の結果、最も差異の小さい標準パターンを認  
5 識結果 1 2 として出力する。パターンマッチングの方法と  
しては、例えば DP マッチングを用いる。

以上説明したとおり、本実施例は騒音下において、「パワ  
ー強度最大の山は背景騒音ではなく、認識対象とする音声  
あるいはその一部である」という仮定に基づいて、

- 10 (1) パワー強度最大の山の区間内には、始端マージン  $bm2$   
と終端マージン  $em1$  以上の端点フリー区間を設定しない
- (2) 端点フリーは、パワー最大の山の跨がない、すなわち  
パワー強度最大の山の後方には、始端フリー区間を含めず、  
パワー強度最大の山の前方には終端フリー区間を含めない  
15 ことに基づいて端点フリー区間を制限するものである。

騒音下では、人間の発声は大きくなることが知られてお  
り、殆どの騒音下で前記仮定が成立する。本実施例による  
音声認識装置は、この仮定に基づくことにより、正確な判  
定が困難な有声音区間の判定を必要とせず、効率的に端点  
20 フリー区間が音声の内部に極力含まれないように制限する  
ことによって、部分マッチングによる誤認識を低減するこ  
とができる。

なお本実施例では、パターンマッチングの方法として DP  
マッチングを用いたが、HMM(Hidden Markov Model)等の  
25 他のパターンマッチング方法に対しても同様の効果が得ら  
れる。また連続音声認識や、不特定話者の単語認識や連続

音声においても同様の効果が得られる。

## 実施例 2 .

次に本発明の別の実施例に係る音声認識装置について説明する。本実施例による音声認識装置では、端点フリー区  
5 間決定手段 6 の動作を以下のように変更する。すなわち、  
まず端点フリー区間決定手段 6 は、入力信号のパワー  $P(t)$   
( $t=1,2,3,\dots,T$ )を入力として、実施例 1 と同様の動作によっ  
て、パワーが予め定めた閾値を越えたフレームが所定数以  
10 上連続する山区間  $PH(i)=\{B(i), E(i)\}$  ( $i=1,2,3,\dots,N$ )と、パワ  
ー強度最大山区間  $PH(I)$ を検出する。次に、パワー強度最  
大山区間  $PH(I)$ 、およびそれに先行する各山区間の始端  
 $B(i)$ , ( $i=1,2,3,\dots,N$ )の前後に始端マージン  $bm1, bm2$ を加え  
た区間を、始端フリー区間  $BF(i)=\{bfL(i), bfR(i)\}$   
15 ( $i=1,2,\dots,I$ )とする。ここで、 $bfL(i)$ は  $i$  番目の始端フリー  
区間の先頭フレームであり、(8)式によって求める。また  
 $bfR(i)$ は  $i$  番目の始端フリー区間の最終フレームであって、  
(9)式によって求める。始端マージン  $bmL(i), bmR(i)$ は、予  
め定めた 0 以上の定数である。

20

$$bfL(i) = B(i) - bmL(i), \quad (i=1, 2, 3, \dots, N) \quad (8)$$

$$bfR(i) = B(i) + bmR(i), \quad (i=1, 2, 3, \dots, N) \quad (9)$$

次に(10)、(11)式にしたがって、パワー強度最大山区  
25 間  $PH(I)$ 、およびそれに後続する各山区間の終端  $E(i)$ ,  
( $i=I, I+1, \dots, N$ )の前後に終端マージン  $em1, em2$ を加えた区

間を、終端フリー区間  $EF(i)=\{efL(i), efR(i)\}$  ( $i=1,2,\dots,N-I+1$ )とする。ここで、 $efL(i)$ は*i*番目の終端フリー区間の先頭フレーム、 $efR(i)$ は*i*番目の終端フリー区間の最終フレームである。始端マージン  $emL(i)$ ,  $emR(i)$ は、予め定めた 0  
 5 以上の定数である。以上の処理により決定された始端フリー区間と終端フリー区間を図3に示す。

$$efL(i) = E(I-1+i) - emL(i), \quad (i=1,2,3,\dots,N-I+1) \quad (10)$$

$$efR(i) = E(I-1+i) + emR(i), \quad (i=1,2,3,\dots,N-I+1) \quad (11)$$

10

なお、前記始端マージン  $bmL(i)$ と  $bmR(i)$ は、始端フリー区間  $BF(i)$ ごとに異なる値に設定してもよいし、共通の値としてもよい。

始端フリー区間を大きくすると、始端の検出誤りには強  
 15 くなるが、部分マッチングの危険性が大きくなるというトレードオフの関係にある。1番目の始端フリー区間  $BF1$ の左側のマージン  $bfL1$ は、音声の外側へのフリー区間であるため、値を大きくしても、部分マッチングの危険性はそれほど増加しない。しかし、1番目の始端フリー区間  $BF1$   
 20 の右側と2番目以降の始端フリー区間は、音声区間内である可能性があり、始端マージン  $bmL(i)$ と  $bmR(i)$ の値を大きく設定すると、部分マッチングの可能性が増す。

そこで、2番目以降の始端フリー区間の左右両側の始端マージンと、1番目の始端フリー区間の右側の始端マージンとを、1番目以降の始端フリー区間の左側の始端マージン  
 25 よりも小さな値若しくは0とすると、内側でのフリー区



間が小さくなる、あるいは 0 となり、部分マッチングを抑える効果がある。

同様に、前記終端マージン  $emL(i)$  と  $emR(i)$  は、終端フリー区間  $EF(i)$  ごとに異なる値に設定してもよいし、共通  
5 の値としてもよい。

最後の終端フリー区間  $EF(N-I+1)$  の右側のマージン  $efR(N-I+1)$  は、音声の外側へのフリー区間であるため、値を大きくしても部分マッチングの危険性はそれほど大きくならない。しかし、最後の終端フリー区間の左側と最後以外  
10 外の終端フリー区間は音声区間内である可能性があるため、終端マージン  $emL(i)$  と  $emR(i)$  の値を大きく設定すると部分マッチングの可能性が増す。

そこで、最後以外の終端フリー区間の左右両側の終端マージンと最後の終端フリー区間の左側とを、最後の終端フリー区間の右側のマージンよりも小さな値若しくは 0 と  
15 することにより、内側でのフリー区間が小さくなる、あるいは 0 となり、部分マッチングを抑える効果がある。

図 4 に、始端マージンを 2 番目以降の始端フリー区間の左右両側の始端マージンと、1 番目の始端フリー区間の右  
20 側の始端マージンを 0 とし、また終端マージンを最後以外の終端フリー区間の左右両側の終端マージンと、最後の終端フリー区間の左側の終端マージンを 0 とした場合を示す。

照合手段 10 は、特徴ベクトルの時系列 4 と、端点フリー区間情報 8 である始端フリー区間  $BF(i) = \{bfL(i), bfR(i)\}$   
25  $(i=1, 2, \dots, I)$  と、終端フリー区間  $EF(i) = \{efL(i), efR(i)\}$   $(i=1, 2, \dots, N-I+1)$  とを入力として、始端フリー区間と終端フ

- リー区間内の全ての始終端の組み合わせにより、標準パターンメモリ 9 に格納されている各単語の標準パターン 10 である REF(i) とパターンマッチングを行う。この処理を標準パターン REF(i), (i=1, 2, 3, ..., K) の全てに対して順次行い、
- 5 距離値が最小のものを認識結果 12 として出力する。パターンマッチングとしては、例えば DP マッチングを用いる。

- 実際の非定常騒音下では、音声の始端と終端でパワーの山の立上りと山の終端が観察されることが多いので、本実施例に係る音声認識装置によれば、実施の形態 1 で説明した
- 10 端点フリー区間の制限に加え、始端のフリー区間はパワーの山の立上り前後の区間、終端のフリー区間はパワーの山の終端の前後の区間に制限することによって、さらに部分マッチングによる誤認識を低減することができる。

#### 15 産業上の利用の可能性

- 本発明は上記のように構成されているので、正確な判定が困難な有声音区間の判定を必要とせず、効率的に端点フリー区間が音声の内部に極力含まれないように制限することによって、部分マッチングによる誤認識を低減すること
- 20 が可能となる。

また本発明は上記のように構成されているので、全ての始終端の組み合わせの中から、標準パターンと最も差異の小さい始終端の組み合わせを選択することが可能となる。

- また本発明は上記のように構成されているので、パワー
- 25 の山の継続時間は短い、瞬間的な信号パワーが大きくなるスパイク状の騒音が発生しうる環境下での音声認識を効

率的に行うことが可能となる。

また本発明は上記のように構成されているので、パワーの山の継続時間は長いが、パワーの極大値は大きくない騒音が発生しうる環境下での音声認識を効率的に行うことが

5 可能となる。

また本発明は上記のように構成されているので、始端と終端の検出誤りを少なくすることが可能となる。

また本発明は上記のように構成されているので、始端と終端と検出誤りを少なくしつつ、部分マッチングの危険性  
10 も抑えることが可能となる。

また本発明は上記のように構成されているので、フレーム単位で音響分析を行う音声認識装置に適用することが可能となる。

## 請 求 の 範 囲

1. 入力音声を音響分析しこの入力信号についてパワーを出力する分析ステップと、
- 5 前記パワーが所定の閾値を連続して超える区間を山区間として検出し、パワーが最大となる山区間を最大山区間として、その最大山区間が前記閾値を下回る点以前に始端フリー区間が存在すると仮定し、前記最大山区間が前記閾値を上回る点以降に終端フリー区間が存在すると仮定して始端
- 10 フリー区間と終端フリー区間の組み合わせを出力する端点フリー区間決定ステップと、  
前記組み合わせの始端フリー区間と終端フリー区間により特定されるパターンと標準パターンとのパターンマッチングを行う照合ステップとを備えたことを特徴とする音声認
- 15 識方法。
2. 前記端点フリー区間決定ステップは、始端フリー区間と終端フリー区間の組み合わせを複数出力し、前記照合ステップは、前記端点フリー区間決定ステップが出力するすべての始端フリー区間と終端フリー区間の組み合わせに
- 20 より特定される各パターンと標準パターンとのパターンマッチングを行う構成とされたことを特徴とする請求の範囲第1項記載の音声認識方法。
3. 前記端点フリー区間決定ステップは、検出された前記山区間のうち、パワーの累積が最大となる山区間を最大
- 25 山区間とする構成とされたことを特徴とする請求の範囲第1項乃至請求の範囲第2項のいずれか一記載の音声認識方

法。

4. 前記分析ステップは、検出点毎にパワーを出力し、  
前記端点フリー区間決定手段は、検出された前記山区間の  
うち、前記検出点毎のパワーのうち上位の所定の個数の検  
5 出点のパワーの和が最大となる山区間を最大山区間とする  
構成とされたことを特徴とする請求の範囲第1項乃至請求  
の範囲第2項のいずれか一記載の音声認識方法。
5. 前記端点フリー区間決定ステップは、前記最大山区  
間以前の山区間が前記閾値を上回る点の近傍に前記始端フ  
10 リー区間が存在すると仮定し、前記最大山区間以後の山区  
間が前記閾値を下回る点の近傍に前記終端フリー区間が存  
在すると仮定する構成とされたことを特徴とする請求の範  
囲第1項乃至請求の範囲第4項のいずれか一記載の音声認  
識方法。
- 15 6. 前記端点フリー区間決定ステップは、前記始端フ  
リー区間が存在すると仮定する範囲の幅と前記終端フリー区  
間が存在すると仮定する範囲の幅とを、各山区間毎に異な  
る構成とされたことを特徴とする請求の範囲第1項乃至請  
求の範囲第4項のいずれか一記載の音声認識方法。
- 20 7. 前記分析ステップは、フレーム毎にパワーを出力し、  
前記端点フリー区間決定手段は、パワーが前記閾値を下回  
るフレームを前記閾値を下回る点とし、パワーが前記閾値  
を上回るフレームを前記閾値を上回る点とする構成とされ  
たことを特徴とする請求の範囲第1項乃至請求の範囲第6  
25 項のいずれか一記載の音声認識方法。
8. 入力音声を音響分析しこの入力信号についてパワー

を出力する分析手段と、

前記パワーが所定の閾値を連続して超える区間を山区間として検出し、パワーが最大となる山区間を最大山区間として、その最大山区間が前記閾値を下回る点以前に始端フリー

- 5 ー区間が存在すると仮定し、前記最大山区間が前記閾値を上回る点以降に終端フリー区間が存在すると仮定して始端フリー区間と終端フリー区間の組み合わせを出力する端点フリー区間決定手段と、

- 10 前記組み合わせの始端フリー区間と終端フリー区間により特定されるパターンと標準パターンとのパターンマッチングを行う照合手段とを備えたことを特徴とする音声認識装置。

9. 前記端点フリー区間決定手段は、始端フリー区間と終端フリー区間の組み合わせを複数出力し、前記照合手段  
15 は、前記端点フリー区間決定手段が出力するすべての始端フリー区間と終端フリー区間の組み合わせにより特定される各パターンと標準パターンとのパターンマッチングを行う構成とされたことを特徴とする請求の範囲第8項記載の音声認識装置。

- 20 10. 前記端点フリー区間決定手段は、検出された前記山区間のうち、パワーの累積が最大となる山区間を最大山区間とする構成とされたことを特徴とする請求の範囲第8項乃至請求の範囲第9項のいずれか一記載の音声認識装置。

11. 前記分析手段は、検出点毎にパワーを出力し、  
25 前記端点フリー区間決定手段は、検出された前記山区間のうち、前記検出点毎のパワーのうち上位の所定の個数の検

出点のパワーの和が最大となる山区間を最大山区間とする構成とされたことを特徴とする請求の範囲第 8 項乃至請求の範囲第 9 項のいずれか一記載の音声認識装置。

- 1 2. 前記端点フリー区間決定手段は、前記最大山区間  
5 以前の山区間が前記閾値を上回る点の近傍に前記始端フリー区間が存在すると仮定し、前記最大山区間以後の山区間が前記閾値を下回る点の近傍に前記終端フリー区間が存在すると仮定する構成とされたことを特徴とする請求の範囲第 8 項乃至請求の範囲第 1 1 項のいずれか一記載の音声認識装置。  
10

- 1 3. 前記端点フリー区間決定手段は、前記始端フリー区間が存在すると仮定する範囲の幅と前記終端フリー区間が存在すると仮定する範囲の幅とを、各山区間毎に異なる構成とされたことを特徴とする請求の範囲第 8 項乃至請求  
15 の範囲第 1 1 項のいずれか一記載の音声認識装置。

- 1 4. 前記分析手段は、フレーム毎にパワーを出力し、前記端点フリー区間決定手段は、パワーが前記閾値を下回るフレームを前記閾値を下回る点とし、パワーが前記閾値を上回るフレームを前記閾値を上回る点とする構成とされたことを特徴とする請求の範囲第 8 項乃至請求の範囲第 1  
20 3 項のいずれか一記載の音声認識装置。

図 1

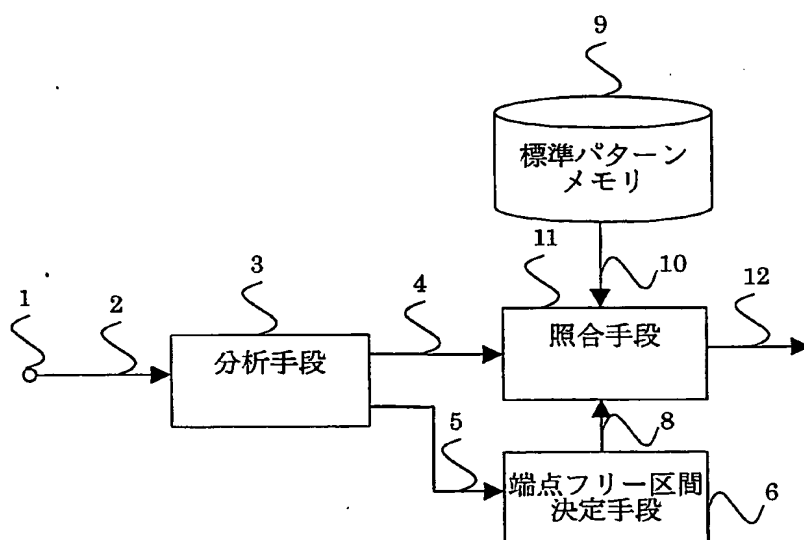




図 2

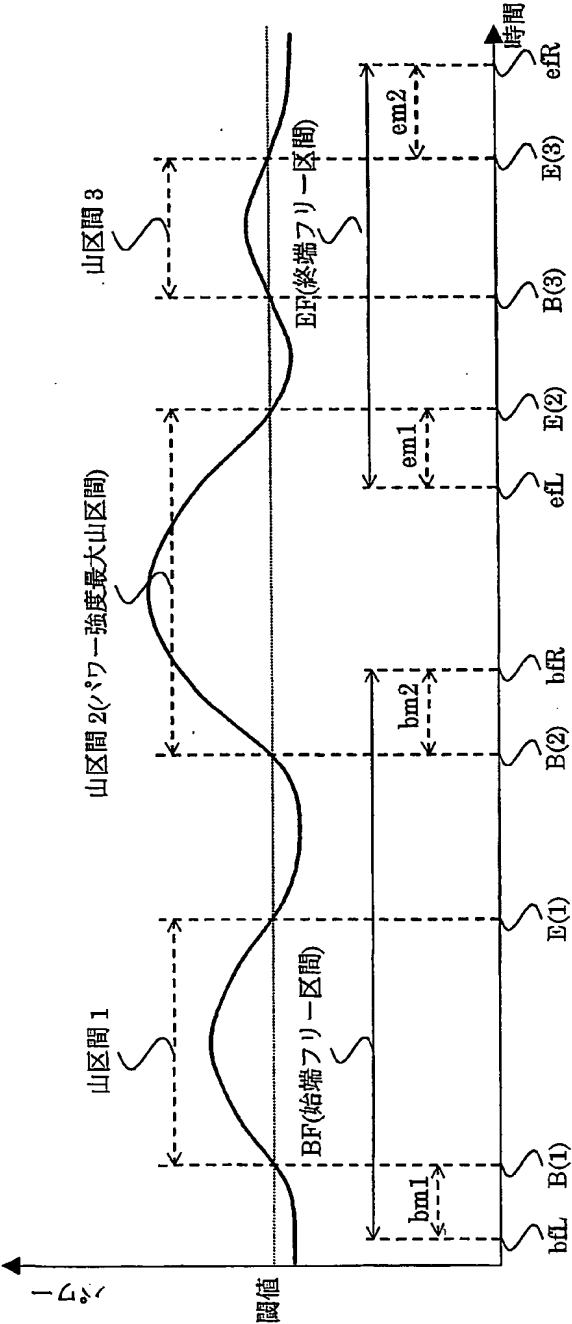


图 3

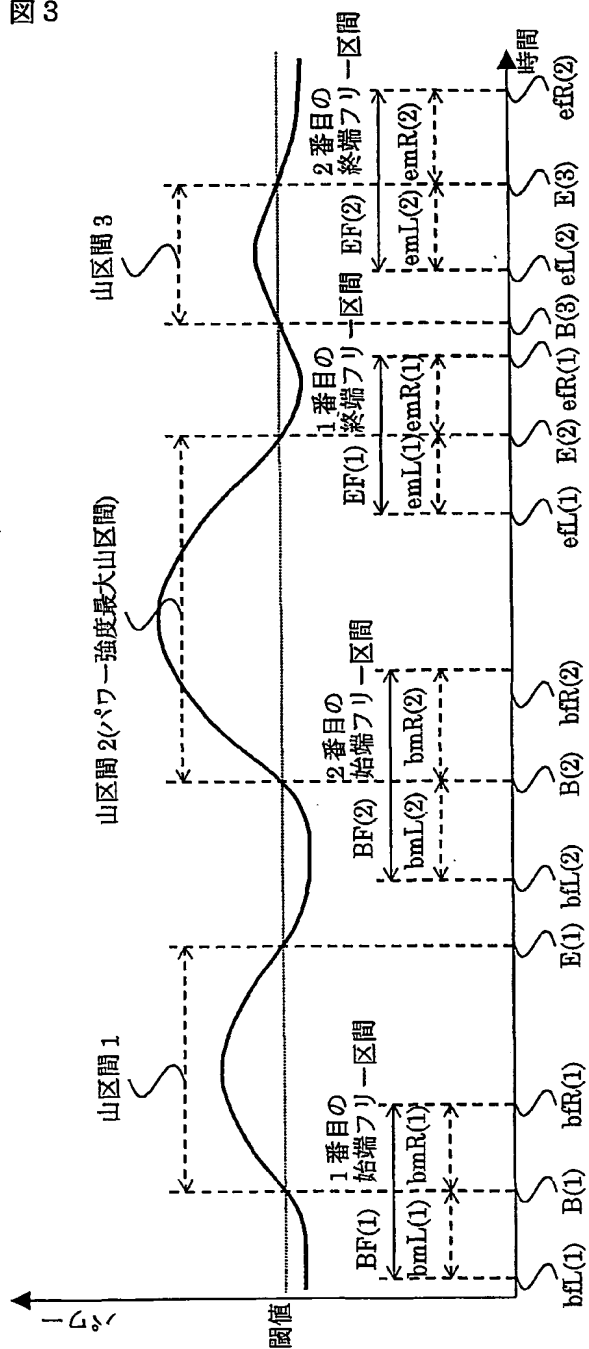


図 4

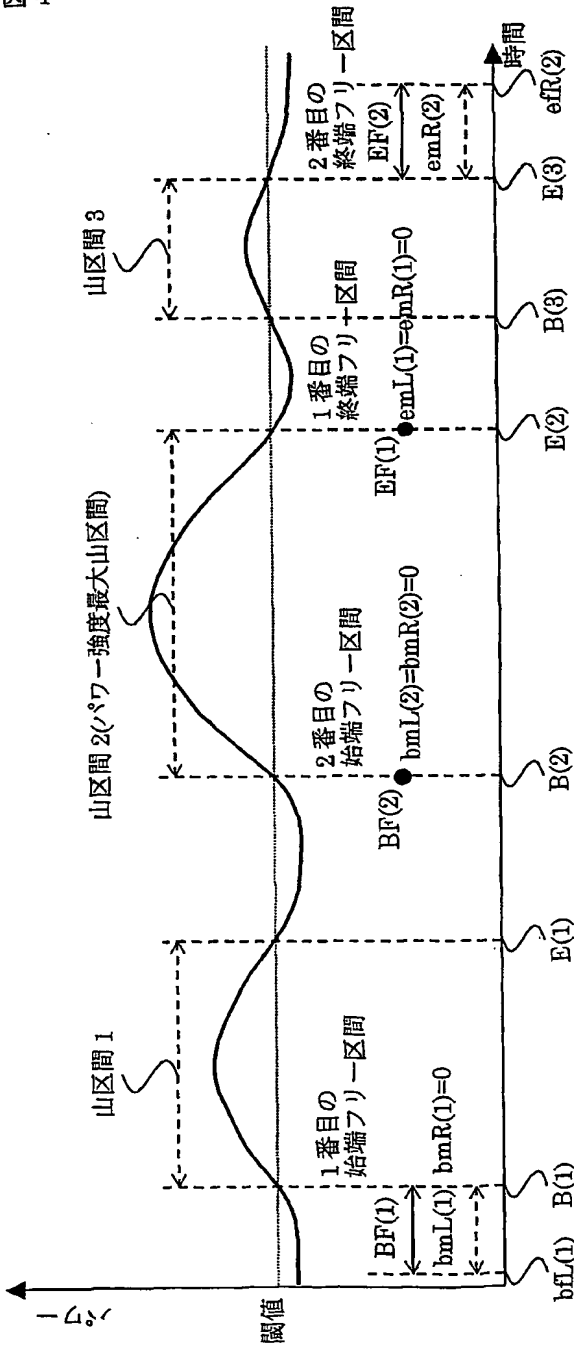


図 5

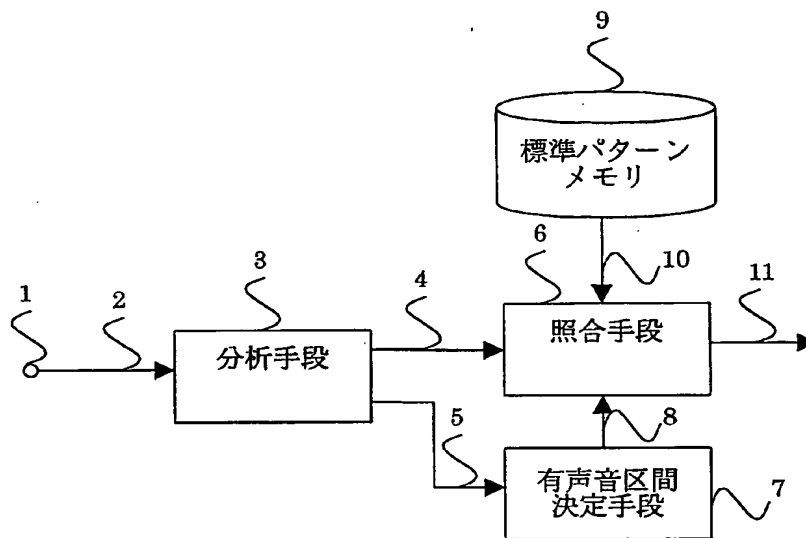
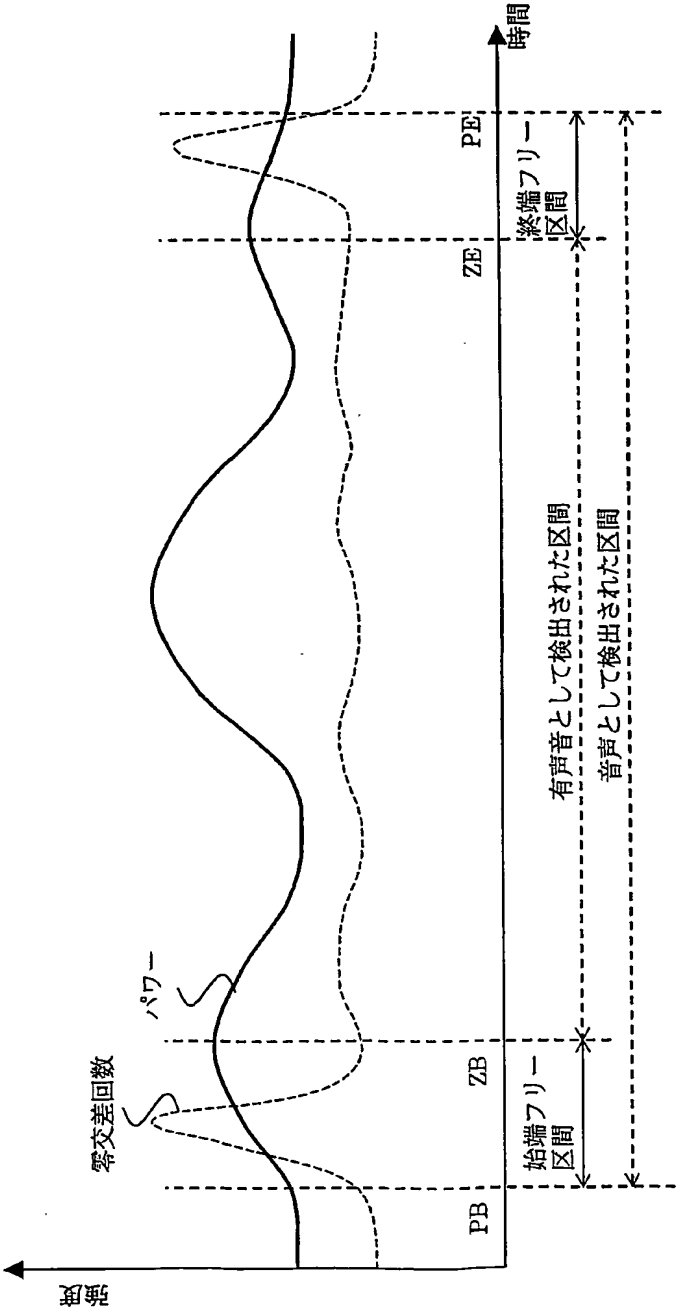


図 6



mm	mm	pp	ppp	eeee	rr	rrr	sssss	oooo	nnnnn		
mmmmmmmm	pp	pp	ee	ee	rrr	rr	ss	oo	oo	nn	nn
mmmmmmmm	pp	pp	eeeeee		rr	rr	ssss	oo	oo	nn	nn
mm	m	mm	ppppp	ee	rr		ss	oo	oo	nn	nn
mm	mm	pp	eeee	rrrr	sssss	oooo	nn	nn			
		pppp									

2222	3333	555555			
22	22	33	33	55	
	22		33	55555	
222	333		55		
22		33	55		
22	22	33	33	55	55
222222	3333		5555		

4/4/05

## 国際調査報告

国際出願番号 PCT/JP02/05847

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G10L11/02

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G10L11/00-11/02

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年

日本国公開実用新案公報 1971~2002年

日本国登録実用新案公報 1994~2002年

日本国実用新案登録公報 1996~2002年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	EP 237934 A (Kabushiki Kaisha Toshiba) 1987. 09. 23, 全文, 全図 & US 4881266 A & DE 3775797 C & JP 6-105394 B	1, 2, 5, 7, 8, 9, 12, 14
Y	JP 2-93696 A (三洋電機株式会社) 1990. 04. 04, 全文, 全図 (ファミリーなし)	1, 2, 5, 7, 8, 9, 12, 14
A		6, 13
A	JP 61-143800 A (株式会社東芝) 1986. 07. 01, 全文, 全図 (ファミリーなし)	1-14

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

05.07.02

国際調査報告の発送日

16.07.02

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

榎本 剛

5C

9379

電話番号 03-3581-1101 内線 3541

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P 2000-250565 A (株式会社リコー) 2000. 09. 14, 全文, 全図 (ファミリーなし)	3, 4, 10, 11
A	J P 8-292787 A (三洋電機株式会社) 1996. 11. 05, 【0084】-【0111】, 第9-11図 (ファミリーなし)	3, 4, 10, 11